National Cancer Institute

eHealth & Availing of the Cloud

Meeting Session

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

NATIONAL CANCER INSTITUTE

caBIG cancer Biomedical Informatics Grid

NCRI INFORMATICS INITIATIVE

NCRI National Cancer Research Institute

Session Aim:

To discuss how cloud computing is being leveraged by caBIG® and the NCI to support the use of applications and resources, using both internal private and public clouds and hosting providers.

Session Participants:

- **William Tulskie**, CEO, CTO, Health Care IT, Inc:
  - "Dr. Susan Love foundation and the Army of Women"
- **Sean Seerey**, Industry Strategist, Microsoft:
  - "Tools for aggregating and presenting cancer outcomes data"
- **Ken Wright and Eric Brinsfield**, SAS Institute Inc.
  - "A Collaborative Approach to Enhancing Patient Outcomes"

Cloud Computing Overview and Applications

R. Mark Adams, Ph.D.
Principal
Booz Allen Hamilton
caBIG® Program Manager

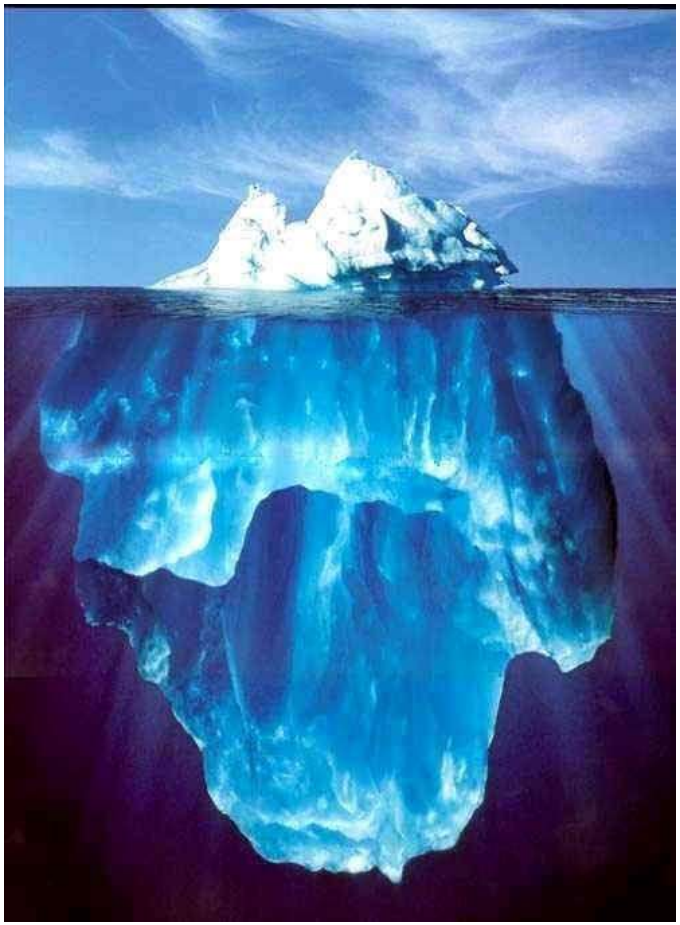
The Cancer Knowledge Cloud
is a virtual biomedical capability that uses the NCI's cancer Biomedical Informatics Grid (caBIG®) to encompass distributed individual and organizational data, software applications, and computational capacity through the Internet.

"Cloud-computing will help to optimize the Federal data facility environment and create a platform to provide services to a broader audience of customers."

President's Budget for FY 2010
Section 9, Cross Cutting Programs

Nearly every federal agency is either moving toward adopting cloud computing or thinking about how to do it

- GSA
- Department of the Interior
  - National Business Center
  - Bureau of Land Management
- NASA
- NOAA
- DISA
- NSA
- FBI
- And on, and on…and on



The issue…

Picture from Chris Dagdigian, BioTeam

Cloud Computing addresses critical issues facing IT

- **Cost:** “Our analysis implies that, over a 13-year life cycle, the total cost of implementing and sustaining a cloud environment may be as much as two-thirds lower than maintaining a traditional, non-virtualized IT data center.” – Booz Allen Hamilton
- **Security**: "Cloud computing can be as secure, if not more secure, than what most organizations do today in the traditional environment“ – Google Enterprise Security Director
- **Information Overload:** Organizations with the largest datasets in the world, use cloud technology to manage and analyze their data.
- **Information Sharing**: “We are at the same point now, in 2010, with Intercloud as we were in ‘73 with Internet” – Vint Cerf, the father of the Internet

Cloud Computing is an evolution of technological models based on proven internet technologies designed to operate at world-scale

Source: IBM

1990

Grid Computing

- Solving large problems with parallel computing

- Made mainstream by Globus Alliance

1998

Utility Computing

- Offering computing resources as a metered service

- Introduced in late 1990s

2000

Software as a Service

- Network based delivery of applications & subscriptions

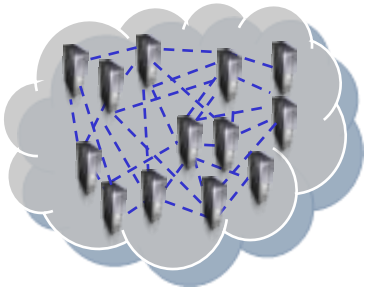
- Gained momentum in early 2000's

2009

Cloud Computing

- Next-generation computing model at world-scale

- Next-generation data centers, business models, unforeseen changes



Scale across all levels of IT operations is the key to Cloud Computing

"Big Data"

- **Scale is the "So-What" factor for Cloud Computing:**
  - The system which can process 10 GB of data can also process 10 PB of data

    Google processes 20 PB of data per day using cloud computing
  - Processing 1000x times more data today is not 1000x harder
  - Column-oriented databases are favored – cloud scale is beyond relational database capacity
  - Provisioning 200 servers is a 3 minute task. Provisioning 2000 servers is a 3 minute task
  - The concept of scale in cloud computing means you always have "enough"
    - Storage
    - Compute
    - Messaging
  - Building IT infrastructure at this scale is expensive - high cost of entry but cloud yields substantial economies of scale
  - Impacts Power, Space, Cooling, COOP, COG

1024 GB = 1 TB
1024 TB = 1 PB
1024 PB = 1 EB

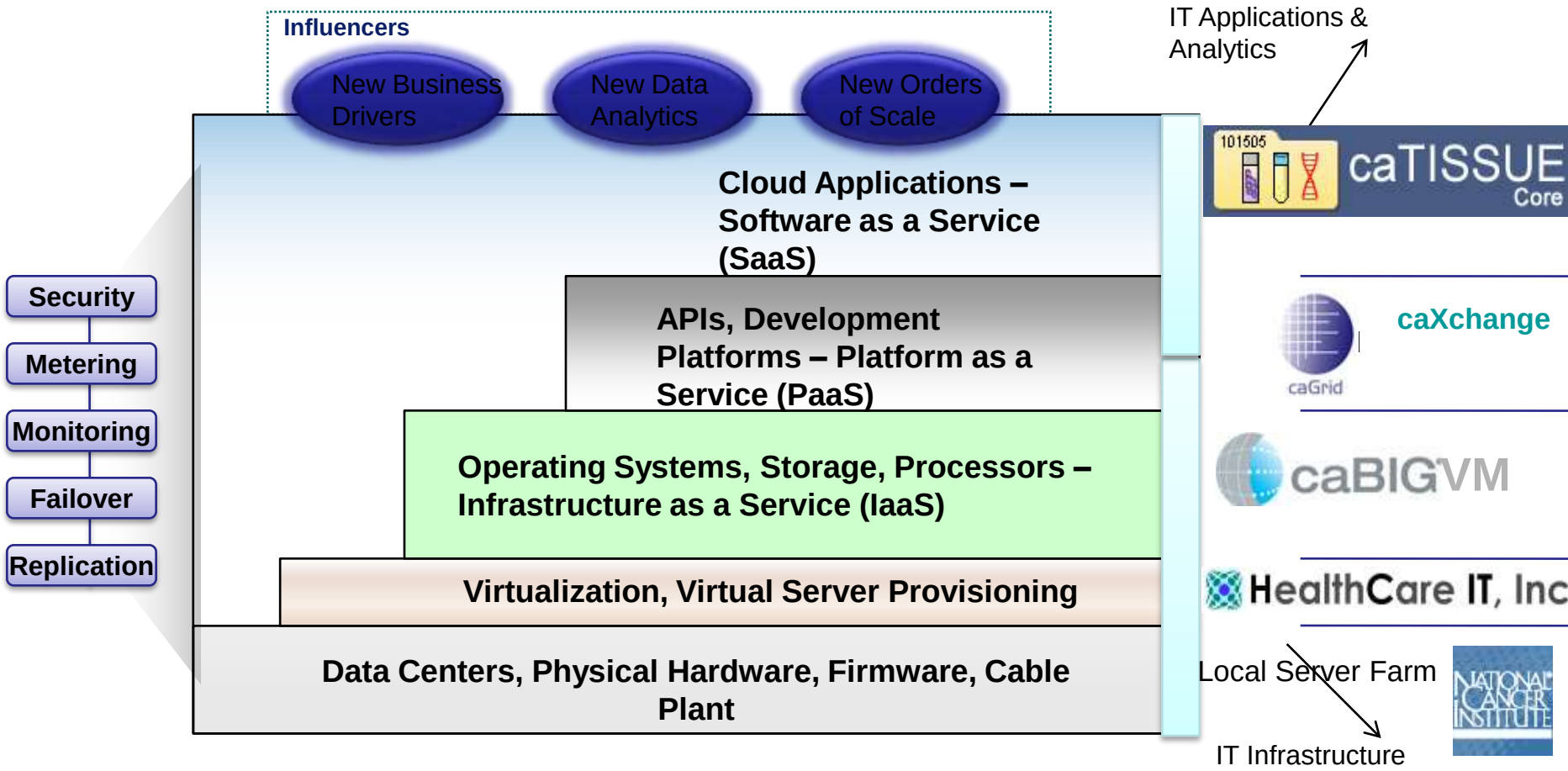
Cloud Foundation: Hardware

- Hundreds or thousands of servers.
- Cloud software is designed to utilize large pools of these servers and must seamlessly handle the removal(machine failures) and addition(more capacity) of machines. This fundamental property is key to cloud's massive scalability.
- Allows applications and corresponding hardware to scale to meet demand.
- Cloud comes in many different sized packages, ***but never as a single server…***

The "Cloud in a Box"
8-20 physical servers

Microsoft Data Center Chicago
2,000 servers per container

Cloud Computing technologies span a range of IT functions & domains to deliver capabilities as a service at world-scale



A working definition of Cloud Computing for us:

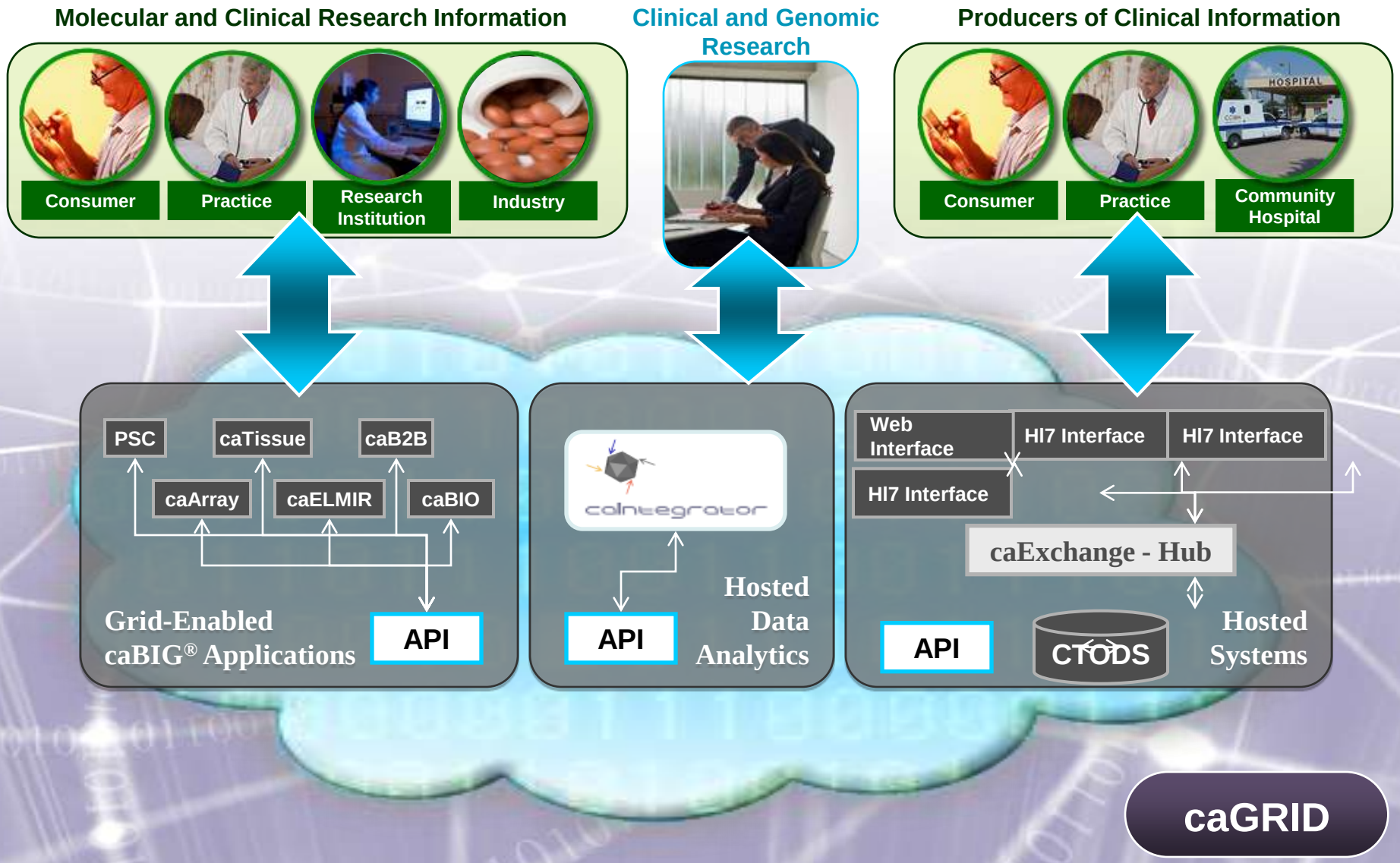
Cloud Computing is…

- A system that can distributed across multiple locations, or run on a single platform
- Tools/infrastructure that can provide access at multiple logical levels to the software and underlying services
- A stack that allows access, management, integration and provisioning of services transparently between systems local, remote and distributed
- Software that can be flexibly accessed from multiple remote locations using installation-free thin clients
- Infrastructure that promotes access via software-as-service to tools and to underlying data via data-as-service to a wide range of users both human and machine

Cloud Computing for the Cancer Knowledge Cloud

- Cloud Computing is…
  - Cloud computing is an approach to computing which provides a means to support software and services running at a range of different locations at multiple levels of service
  - Cloud computing provides an avenue for smoothly transitioning to more distributed and less local infrastructure under demand
- For the Cancer Knowledge Cloud…
  - Cloud computing will need to provide flexible accommodation for varying Federal, state and local regulations regarding the location and transfer of the underlying data
  - Cloud computing security approaches will need to be flexible in adjusting to the varying nature and access needs of a wide range of biomedical data types with varying privacy, ethical, regulatory and IP requirements
  - Access to cloud computing services will need to occur at a range of different logical levels, ranging from direct access to underlying infrastructure features through APIs at a range of levels to end-user interfaces

IT-Enabled Research



Cloud Computing is not a panacea for IT – some capabilities are not right for the cloud

- Cloud Computing Readiness Assessment & Transition Models – not everything belongs in the cloud:
  - Real-time applications, Command & Control systems are probably not cloud candidates
- Legal & Regulatory Challenges:
  - HIPPA, SOX, privacy issues with physical data located outside US
- Multi-Tennant Environment:
  - All resources are shared (disk storage, network) causes difficulty cleaning 'spills'
- Security
  - Can security actually be *increased* by moving to a cloud?
  - What new security vulnerabilities are introduced?
- Maturity
  - Cloud computing is proven at Google, Microsoft, Amazon, using proprietary technology –the open-source cloud technologies continue to mature

Facing the Challenges Around Implementing Clouds

- **Security & Trust** – There are concerns about loss of control and vulnerabilities. However, in many ways Cloud Computing may offer security and trust models that are improvements on traditional infrastructure approaches
- **Maturity** – The full suite of technology enablers, including development and management tools, are still in early stages of development. So too are policies and business models required for a full scale transition to Cloud Computing
- **Education & Change Management** – Cloud computing is emerging and is not well known or understood. There are multiple vectors of education required to enable mass adoption of cloud computing, e.g. what types of applications are most enabled by the cloud, what are the transition risks, best practice and process definitions
- **Service Levels & Performance Perceptions** – It is unclear at this point the service level agreements that will be required from the Cloud Computing providers; it is further uncertain and unknown the type of certifications the Government will request from these providers before they place their critical data and mission driven applications in their infrastructure. There are also likely to be issues around performance and other service areas
- **Transition** – Significant challenges exist to help transition the Enterprise or part of the Enterprise to the Cloud

Session Participants:

- **William Tulskie**, CEO, CTO, Health Care IT, Inc:
  - "Dr. Susan Love foundation and the Army of Women"
- **Sean Seerey**, Industry Strategist, Microsoft:
  - "Tools for aggregating and presenting cancer outcomes data"
- **Ken Wright and Eric Brinsfield**, SAS Institute Inc.
  - "A Collaborative Approach to Enhancing Patient Outcomes"